Robustness of shape descriptors to incomplete contour representations

Ghosh, Anarta; Petkov, Nicolai

Published in:
Ieee transactions on pattern analysis and machine intelligence

Publication date:
2005

Robustness of Shape Descriptors to Incomplete Contour Representations

Anarta Ghosh and Nicolai Petkov

Abstract—With inspiration from psychophysical researches of the human visual system, we propose a novel aspect and a method for performance evaluation of contour-based shape recognition algorithms regarding their robustness to incompleteness of contours. We use complete contour representations of objects as a reference (training) set. Incomplete contour representations of the same objects are used as a test set. The performance of an algorithm is reported using the recognition rate as a function of the percentage of contour retained. We call this evaluation procedure the ICR test. We consider three types of contour incompleteness, viz. segment-wise contour deletion, occlusion, and random pixel depletion. As an illustration, the robustness of two shape recognition algorithms to contour incompleteness is evaluated. These algorithms use a shape context and a distance multiset as local shape descriptors. Qualitatively, both algorithms mimic human visual perception in the sense that recognition performance monotonously increases with the degree of completeness and that they perform best in the case of random depletion and worst in the case of occluded contours. The distance multiset method performs better than the shape context method in this test framework.

Index Terms—Contour, COIL, deletion, depletion, distance multiset, Gollin, incompleteness, ICR test, MPEG-7, object recognition, occlusion, psychophysics, shape, shape context.

1 Introduction

If we look at the objects in Fig. 1, we can instantly recognize birds, even though 50 percent of the contour is removed segment-wise in Fig. 1a, the right half of the contour is not visible in Fig. 1b, and 80 percent of the contour points have been removed (randomly) in Fig. 1c. This ability of human beings to recognize objects with incomplete contours was studied by psychologist E.S. Gollin [1]. His objective was to investigate the performance of humans in recognizing objects with incomplete contours as a function of developmental characteristics, such as mental and chronological age and intelligence quotient. The subjects of his experiments were children of different age groups and a group of adults. Gollin used sets of contour[1] images with different degrees of incompleteness (Fig. 2) and addressed the following questions:

1. In order to be recognized, how complete do the contours of common objects need to be?
2. How does training affect the recognition performance in case of incomplete representations?

The main conclusions drawn by him through his experiments were: human ability to recognize objects with incomplete contours

a. depends on intelligence quotient and
b. is improved by training.

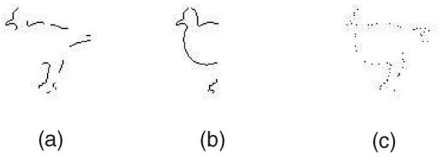
In the context of processing visual information using computers, this aspect of recognition of objects with incomplete contours is also very important. Fig. 3 shows a natural image and two edge images obtained from it. Fig. 3b was obtained by applying a bank of Gabor energy filters. It contains the contours of the object of interest, viz. a rhinoceros, but it also contains a large number of texture edges in the background that are not related in any way to the shape of the rhinoceros. These texture edges will have a devastating effect on the performance of all currently known contour-based shape recognition algorithms. Advanced contour detection methods based on surround suppression [2], [3] succeed in separating the essential object contours from the texture edges, as illustrated by Fig. 3c, but, at the same time, these methods have a certain negative side effect of depleting the contours of the objects of interest. Hence, the robustness of shape recognition methods to contour incompleteness is an issue of practical importance.

With inspiration from Gollin's work, we propose a novel attribute, viz. *robustness to incomplete contour representations*, that any contour-based object recognition system/algorithm should have. The objective of this study is to show how the performance of recognition systems/algorithms can be investigated in an idealized situation where: 1) Complete contour representations of the objects to be recognized form the reference (training) set or "memory" of the system/algorithm, 2) incomplete contour representations of the same objects are derived from the aforementioned complete representations and are used as a test set, and 3) the performance of the system/algorithm in recognizing the objects from these incomplete representations is evaluated. The main reason behind choosing such an ideal situation is the rational logic that, in order to perform well in a real-world scenario (natural images), any recognition system should first perform well in such idealized (simple) situations. In his study [1], Gollin also worked in a similar idealized situation

1. By "contours" in the following, we refer to both occluding boundaries and inner edges that are defined by boundaries of parts of an object or perceptually important color or texture regions.

• *The authors are with the Institute of Mathematics and Computing Science, University of Groningen, PO Box 800, 9700 AV, Groningen, The Netherlands. E-mail: {anarta, petkov}@cs.rug.nl.*

Manuscript received 9 July 2004; revised 3 Feb. 2005; accepted 16 Feb. 2005; published online 14 Sept. 2005.
Recommended for acceptance by R. Basri.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0342-0704.



(a) (b) (c)

Fig. 1. A bird can be recognized even though (a) 50 percent of its contour has been removed segment-wise, (b) its tail part is not visible (occluded), and (c) 80 percent of the contour pixels have been randomly removed.

where it was assumed that the subjects are familiar with the complete representations of the objects to be recognized. To be precise, if any of the subjects could not recognize the complete representations of any of the objects, then his data were discarded from the study.

In this paper, we study the robustness of contour-based shape recognition methods by comparing an object represented by incomplete contours with all objects in a reference set represented by complete contours and determining the nearest neighbor. If the nearest neighbor is the object from which the incomplete contour representation is derived, we consider the recognition to be correct, otherwise incorrect. We also propose possible extensions and generalizations of this basic framework.

In addition to Gollin's method of segment-wise contour deletion (like Set I to Set IV of Fig. 2), we also consider other types of incompleteness, viz. occlusion and random pixel depletion. We name the corresponding studies segment-wise deletion test, occlusion test, and depletion test. Collectively, we call these tests, in short, *Incomplete Contour Representations (ICR) tests*.

The choice of the shape recognition methods we study is limited by the condition that they use contour information. Unless necessary modifications are done, methods which use other type of information fall outside the scope of this study. For instance, Gavrila [4] proposes a method based on the

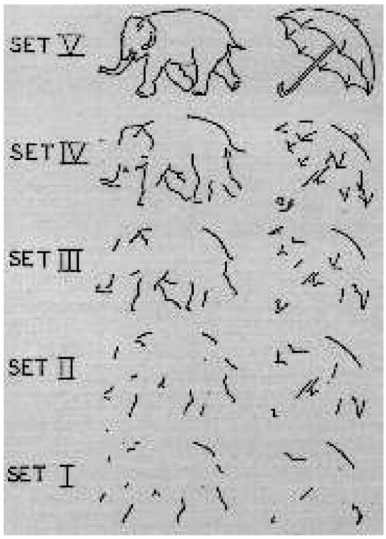
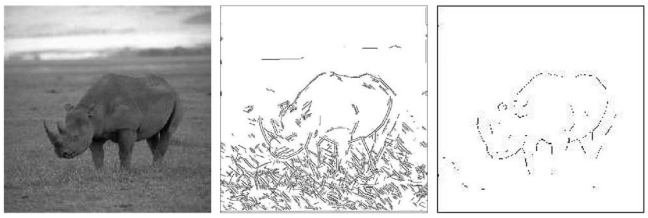


Fig. 2. Example of image sets used in Gollin's original test [1]. The images in Set V are complete contour representations and the other sets are derived from Set V by removing segment-wise an increasing fraction of the contour. Reproduced with permission of the author and publisher of: E.S. Gollin, "Developmental Studies of Visual Recongition of Incomplete Objects," *Perceptual and Motor Skills*, vol. 11, pp. 289-298, Southern University Press, 1960.



(a) (b) (c)

Fig. 3. (a) Image of a rhinoceros in its natural habitat. (b) Result of edge detection with a bank of Gabor energy filters. (c) Result of contour detection by a bank of Gabor energy filters augmented with a biologically motivated surround suppression of texture edges. The contours of the object of interest are more visible in the latter image, but the suppression of texture edges has resulted in a partial contour depletion. (Both algorithms are available as Web applets at www.cs.rug.nl/~petkov.)

distance transform in which every point of a binary object is characterized by its distance to the object's border. In our study, objects are represented by their contour points only and, hence, the distance transform is not informative. Due to the same reason, Goshtaby's shape matrix [5] cannot be directly assessed in this framework either. Some other methods, which do not use boundary information are the medial axis transform approach described in [6], [7], [8] and the moment-based approach dealt with in [9], [10]. Latecki and Lakämper's polygonal shape descriptor [11] inherently assumes that an object is represented by a closed curve and, therefore, this method first needs some modification before it can be applied to objects represented by incomplete contours. Mokhtarian and Mackworth's curvature scale space method [12], which plays an important role in the MPEG-7 standard, computes the curvature at every point of a closed curve (at different scales) to represent the shape of an object. Hence, this method also needs some modification (e.g., estimation of the missing portions of the contours) in order to be evaluated in the ICR test framework. For further aspects of and references to shape analysis and object recognition methods see, e.g., [13], [14], [15], [16], [17], [18], [19].

In this paper, we study the *shape context* method described in [20] and the *distance multiset* method described in [13] with respect to their robustness to contour incompleteness of different types. In Section 2, we briefly describe these methods. In Section 3, we present the basic experimental design and the achieved results. We discuss some further aspects and possible test extensions in Section 4. A summary and conclusions are presented in Section 5.

2 Contour-Based Shape Recognition Methods

In both methods studied below, the recognition of objects is done by computing dissimilarity between the contour representations of two objects by using a point correspondence paradigm. The point correspondences are found using shape descriptors associated with the points.

2.1 Shape Context

A shape descriptor, called the *shape context* [20], of a point p belonging to the contour of an object is a bivariate histogram in a log-polar coordinate system that gives the distribution of contour points in the surroundings of p. Let an object $\mathcal{O}$ be represented by a set of contour points, $\mathcal{O} \equiv \{p_1 \dots p_N\}$.

Formally, the authors of this method define the shape context of a point $p \in \mathcal{O}$ as a vector in the following way:

$$H_K^{\mathcal{O}}(p) = (h_1(p), h_2(p), \ldots, h_K(p)), \quad (1)$$

where

$$h_k(p) = card\{q \neq p | q \in \mathcal{O}, (q-p) \in bin(k)\} \quad (2)$$

is the number of contour points in the kth bin $bin(k)$ and K is the total number of histogram bins. The bins are constructed by dividing the image plane into K partitions (in a log-polar coordinate system) with p as the origin. In this study, we use five intervals for the log distance r and 12 intervals for the polar angle θ, so $K = 60$. As the radius of the surroundings on which the shape context is computed (the upper bound of the radial distance r) we choose the diagonal of the image. In this way, this radius is constant for all experiments. As suggested in [20], we randomly choose 100 points (if available) from the contour of an object and calculate their shape contexts. The shape of the object is described using the set of shape contexts associated with the contour points in the following way:

$$S_{\mathcal{O}}^{SC} \equiv \left\{ H_K^{\mathcal{O}}(p) | p \in \mathcal{O} \right\}. \quad (3)$$

The cost of matching a point p_i that belongs to the contour of an object $\mathcal{O}_1$ of M points, to a point q_j from the contour of an object $\mathcal{O}_2$ of N points is defined as follows:

$$c_{i,j}^{SC} \equiv \frac{1}{2} \sum_{k=1}^{K} \frac{[h_k(p_i) - h_k(q_j)]^2}{h_k(p_i) + h_k(q_j)}. \quad (4)$$

An $M \times N$ cost matrix of point-wise dissimilarities is constructed according to (4). Next, we compute the dissimilarity between the shapes $S_{\mathcal{O}_1}^{SC}$ and $S_{\mathcal{O}_2}^{SC}$ of the objects in the following way:

$$d^{SC}\left(S_{\mathcal{O}_1}^{SC}, S_{\mathcal{O}_2}^{SC}\right) \equiv \sum_{i=1}^{M} min\left\{ c_{i,j}^{SC} | j = 1, \ldots, N \right\}. \quad (5)$$

The authors of the shape context approach [20] (and also the authors of the distance multiset approach [13]) use a different method to compute the dissimilarity of two shapes from the point-wise dissimilarity matrix. More specifically, they use the Hungarian algorithm [21] of bipartite graph matching to solve the optimal assignment problem. In our experiments, we found that the simple method according to relation (5) gives sufficient[2] results. Further aspects of the shape context method as presented in [20], such as a thin plate spline transform and certain sampling considerations are not deployed here for simplicity and in order to make the two algorithms used comparable regarding the number and complexity of processing steps.

2.2 Distance Multiset

For a point p in the contour of an object $\mathcal{O}$ of N points, the *distance multiset*[3] is formally defined as the following vector [13]:

2. By sufficient, we mean that the recognition rate is good enough to illustrate the conceptual aspects of the ICR test framework. We use the same simple method to compute the dissimilarity of two shapes for both algorithms that are studied.

3. In [13], the term distance *set* is used, which is not always correct since this data structure might contain repeating elements and should therefore be called *multiset* or *bag*.

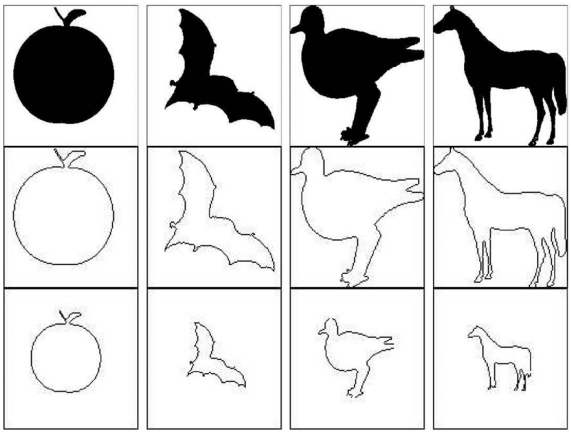


Fig. 4. Row 1: Sample of the MPEG-7 data set [22] images used. Row 2: Contour images extracted from the images in Row 1. Row 3: Rescaled contour images, these images are considered as complete representations that comprise the memory of the recognition system.

$$D_N^{\mathcal{O}}(p) = (ln(d_1(p)), ln(d_2(p)), \ldots, ln(d_{N-1}(p))), \quad (6)$$

where $d_j(p)$ is the Euclidean distance between p and its jth nearest neighbor in $\mathcal{O}$. In this approach, the shape of an object $\mathcal{O} \equiv \{p_1 \ldots p_N\}$ defined by a set of contour points is described by the set of distance multisets in the following way:

$$S_{\mathcal{O}}^{DM} \equiv \left\{ D_N^{\mathcal{O}}(p) | p \in \mathcal{O} \right\}. \quad (7)$$

Next, a cost $c(X, Y)$ of matching two distance multisets X and Y is defined, see Appendix A.

Let $c_{i,j}^{DM}$ be the cost of matching a point p_i in an object $\mathcal{O}_1$ represented by M contour points to a point q_j in an object $\mathcal{O}_2$ represented by N contour points, $M \leq N$:

$$c_{i,j}^{DM} \equiv c(D_N^{\mathcal{O}_1}(p_i), D_M^{\mathcal{O}_2}(q_j)). \quad (8)$$

Similar to (5), the dissimilarity between the shapes $S_{\mathcal{O}_1}^{DM}$ and $S_{\mathcal{O}_2}^{DM}$ is defined as follows:

$$d^{DM}\left(S_{\mathcal{O}_1}^{DM}, S_{\mathcal{O}_2}^{DM}\right) \equiv \sum_{i=1}^{M} min\left\{ c_{i,j}^{DM} | j = 1 \ldots N \right\}. \quad (9)$$

Further aspects of the distance multiset method as presented in [13], such as the use of multiple features in a data structure called the labeled distance set, are not deployed here for the reasons mentioned at the end of Section 2.1.

3 Experiments and Results

3.1 Data Set

As a data set, we choose images from the MPEG-7 data set [22]. It contains 1,400 images divided into 70 classes, each of 20 similar objects (e.g., apple, bird, bat, etc). We choose one object from each class (Fig. 4, Row 1) and extract the contours of the object using Gabor filters [3] (Fig. 4, Row 2). The resulting 70 contour images are rescaled in such a way that the diameter (maximum Euclidean distance between two contour pixels) is approximately the same (76 pixels) for all objects, cf., Row 3 of Fig. 4. These 70 rescaled contour images are used as reference images in our experiments. The set of these images corresponds to the complete representations, Set V of Fig. 2, used in Gollin's original study and form the "memory" of the recognition system.

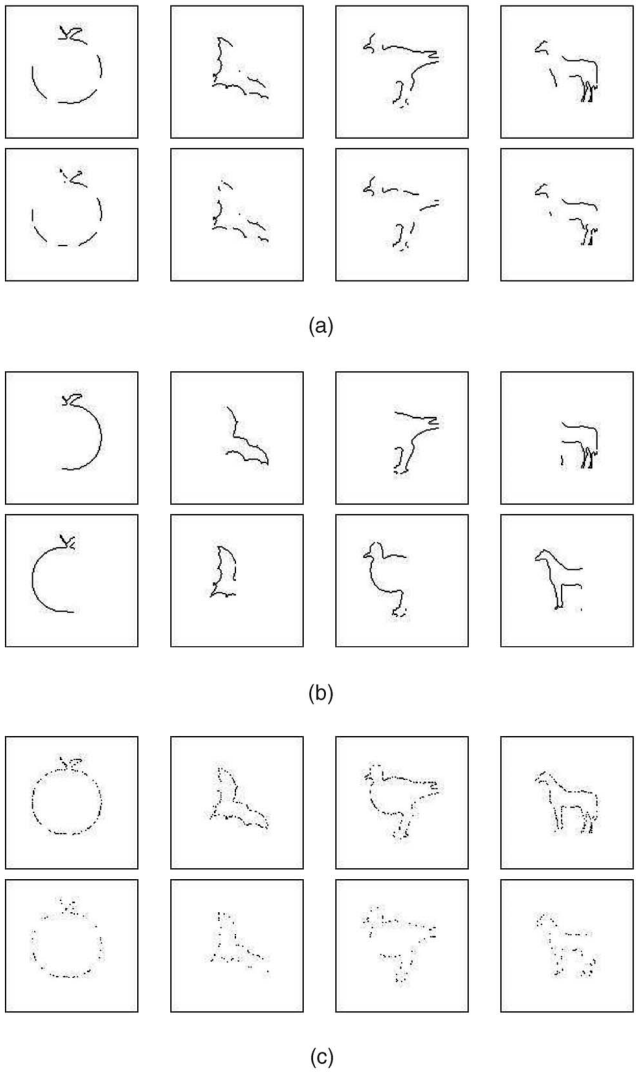


Fig. 5. (a) Segment-wise deleted contour representations of objects (they correspond to the incomplete representations of Gollin's original study, Set I to IV of Fig. 2). The images are obtained by retaining 70 percent (Row 1) and 50 percent (Row 2) of the contour pixels (compare with Row 3 of Fig. 4). (b) Occluded contour representations of objects. The images are obtained by removing 40 percent of the contour from left (Row 1) and right (Row 2). (c) Depleted contour representations of objects which are constructed by randomly removing 50 percent (Row 1) and 80 percent (Row 2) of the contour pixels.

For the *segment-wise deletion test*, incomplete representations (Fig. 5a) are constructed by randomly removing continuous segments of the contours and retaining a given percentage of contour pixels from the above mentioned complete contour representations. We delete approximately $\lceil log_2(\frac{100-c}{8}) \rceil$ segments for percentage c of retained pixels.

For the *occlusion test*, incomplete representations are created by removing a given percentage of consecutive contour pixels starting from the leftmost (Row 1 of Fig. 5b) or the rightmost pixel (Row 2 of Fig. 5b) of an object. The choice of left and right occlusion in our study is driven by the fact that, in the case of natural images, the object of interest is most commonly occluded either from the left or from the right.

For the *depletion test*, the incomplete representations (Fig. 5c) are obtained by randomly removing a given percentage of pixels from the contours of the complete contour representations.

In our experiments, the percentages of retained pixels are chosen in the following way: from 2 to 4 percent in steps of 1 percent, from 5 to 85 percent in steps of 5 percent, and 100 percent for the depletion test; from 5 to 85 percent in steps of 5 percent, and 100 percent for the segment-wise deletion and the occlusion tests. For each type (segment-wise deletion, occlusion, and depletion) and degree of contour degradation, we create 70 test images from the corresponding reference images. All complete contour images and incomplete contour images obtained with different types and percentages of incompleteness are available at www.cs.rug.nl/~petkov.

3.2 Method

An incomplete representation (segment-wise deleted or depleted or occluded contour image) obtained from one of the 70 reference images is compared with all 70 reference images and a decision is made about which reference image the degraded image is most similar to (nearest neighbor search). The comparison is based on a shape dissimilarity computed using a given shape comparison algorithm, described in Section 2. If the nearest neighbor is the reference image from which the degraded image was obtained, the recognition is considered correct, otherwise incorrect. If the nearest neighbor is found to be not unique, then the recognition is also considered incorrect. For each of the three tests (segment-wise deletion, occlusion, depletion) and for each degree of contour image degradation, the corresponding 70 test images are compared with each of the 70 reference images and the percentage of correct recognition is determined. The percentage of correct recognition P is observed as a function $P(c)$ of the percentage c of retained contour pixels. In the case of the occlusion test, the percentage of correct recognition is calculated by averaging the correct recognition rates with the left and right occluded images for a given percentage of retained contour.

3.3 Results

Fig. 6 shows the results of our experiments. In all three tests and for both shape comparison algorithms, the recognition rate is a monotonously increasing function of the percentage of contour retention. In this respect, the considered algorithms resemble the human visual system [23], [24], [25]. Both methods perform the worst in the occlusion test and the best in the depletion test, which also conforms with the recognition performance of humans as occluded contour images carry the least amount of shape information and depleted contour images carry maximum shape information in the context of human visual perception.

In the case of the segment-wise deletion test (Fig. 6a) and the occlusion test (Fig. 6b), the performance of the distance multiset method is appreciably better than that of the shape context method for any percentage of retained contour pixels. From the results of the depletion test (Fig. 6c), we see that both the shape context method and the distance multiset method perform very well in recognizing objects with depleted contour representations if more than 40 percent and 5 percent, respectively, of the contour points are retained. The distance multiset method outperforms the shape context method when the degree of depletion is very high, i.e., a very low percentage (less than 40 percent) of the pixels are retained.

For both methods, the results of the occlusion test are worse than the results in other tests. This is more evident for the shape context method and can be explained as follows:

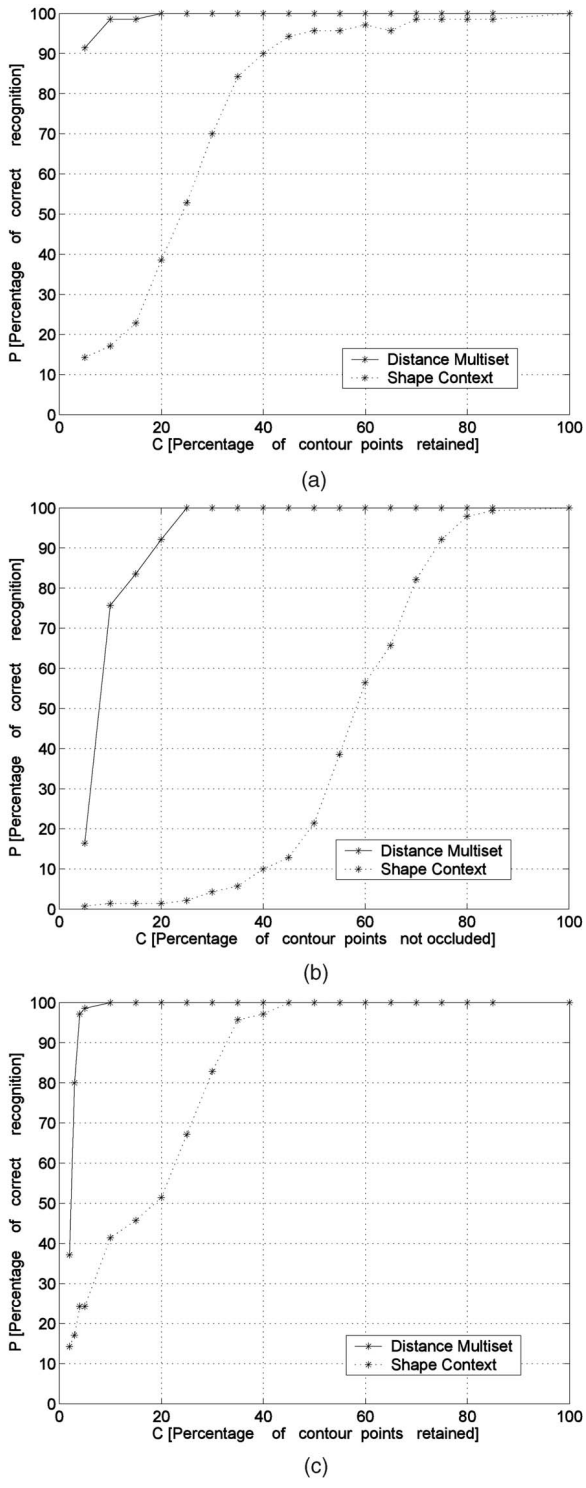


Fig. 6. Results of the ICR tests with a subset of MPEG-7 data set: (a) Segment-wise deletion test: The distance multiset method gives 100 percent recognition rate when more than 20 percent of the contour is retained. The shape context method performs reasonably well (more than 90 percent recognition rate) when more than 40 percent of the contour points are retained. (b) Occlusion test: The shape context method is particularly affected by occlusion because the shape context descriptors of contour points near the occlusion boundary are radically different from the shape contexts of the points in the complete contours. (c) Depletion test: The distance multiset method and the shape context method perform very well for more than 5 percent and 40 percent retention of contour pixels.

A contour point near the occlusion boundary has radically different shape context from the same contour point in the reference (unoccluded) object contour since all the contour points on one side of the occlusion boundary are missing. Hence, such a point will make a large contribution to the dissimilarity between occluded and unoccluded contours.

In general, the better performance of the distance multiset method can be explained by the fact that the proposed ICR tests give an advantage to algorithms which yield zero dissimilarity in a comparison of two objects represented by two sets of points where one is a subset of the other. This property of the distance multiset algorithm is explained in more detail below. (Another method with the same property is based on the nonsymmetric Hausdorff distance.) Let us consider two sets $A, B, \subset \mathbf{R}^2$ such that

$$B = \{f(\mathbf{x}) : \mathbf{x} \in A\}, \tag{10}$$

where $f : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ is defined as follows:

$$f(\mathbf{x}) = \mathbf{L}\mathbf{x} + \mathbf{t}, \forall \mathbf{x} \in \mathbf{R}^2, \tag{11}$$

$\mathbf{L}$ being a 2×2 orthogonal matrix ($|det(\mathbf{L})| = 1$) and $\mathbf{t} \in \mathbf{R}^2$. Note that the transformation (11) preserves the Euclidean distance between points (isometry). The following special forms of f are of particular interest:

1. $\mathbf{L} = I$ (the identity matrix) and $\mathbf{t} = \mathbf{0}$: identity transformation, $B = A$.
2. $\mathbf{L} = I$ and $\mathbf{t} \neq \mathbf{0}$: pure translation, B is a translated version of A.
3. $det(\mathbf{L}) = 1$, $\mathbf{t} = \mathbf{0}$: pure rotation, B is a rotated version of A.
4. $det(\mathbf{L}) = -1$, $\mathbf{t} = \mathbf{0}$: pure reflection, the elements of B, are obtained by reflecting elements of A across a straight line.

So, if A is the set of contour points of an object $\mathcal{O}_1$, then B is the set of contour points of an object $\mathcal{O}_2$ that is derived from $\mathcal{O}_1$ through any of the transformations described in items 1 through 4 or any combination thereof.

Lemma. *Let B be obtained from A according to (10)-(11) and let C be a subset of B,*

$$C \subset B, \ \mathrm{card}(C) \geq 2. \tag{12}$$

It holds

$$d^{DM}\left(S_C^{DM}, S_A^{DM}\right) = 0, \tag{13}$$

where S_C^{DM} and S_A^{DM} are the shapes, described by distance multisets, corresponding to C and A, respectively.

The proof of this lemma is included in Appendix B at the end of the paper. In our study, A corresponds to the set of contour points of a reference object, f is the identity transformation (i.e., $B = A$), and C is the set of contour points of an incomplete representation.

The implication of the lemma is two-fold:

1. In the case of the distance multiset method, the recognition will be incorrect only when the nearest neighbor of a test object in the reference set is not unique.

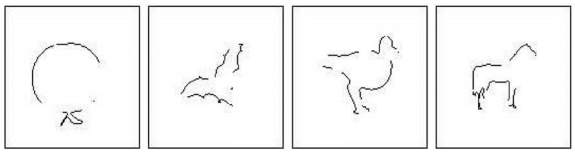


Fig. 7. Sample of incomplete contour representations of affine transformed objects used in a segment-wise deletion test (compare with rows of Fig. 4).

2. The distance multiset method should perform exactly the same way when f is not the identity transformation, that is, the incomplete representations of the objects are not derived directly from the reference objects but from affine transformed versions of them. This emphasizes the use of a distance multiset in our study instead of just matching pixels to calculate the dissimilarity between the shapes of objects.

To illustrate the latter implication 2, we consider the incomplete representations depicted in Fig. 7. Here, the incomplete representations are not directly derived from the objects depicted in Row 3 of Fig. 4 but from affine transformed versions of them. The affine transformations are chosen to be either rotation (by angle $n\frac{\pi}{2}$ to avoid discretization effects) or reflection across a line. We performed the segment-wise deletion test on the distance multiset method using incomplete representations such as those shown in Fig. 7. The results of this experiment are shown in Fig. 8. If we compare the original segment-wise deletion test results for the distance multiset method (Fig. 6a) with these results, we do not find any qualitative difference. A quantitative difference might arise due to the randomness involved in the construction of incomplete representations.

The above lemma does not hold for the shape context method, but this method can be modified in such a way that the relation (13) can be approximately fulfilled. Specifically, we normalize the shape context $H_K^{\mathcal{O}}(p)$ by dividing its elements by the total number of points $card(\mathcal{O})$ in the corresponding object $\mathcal{O}$. If $\mathcal{O}' \subset \mathcal{O}$ is an incomplete representation derived from $\mathcal{O}$ and $H_K^{\mathcal{O}'}(p)$ is the normalized (by $card(\mathcal{O}')$) shape context of a point $p\,(p \in \mathcal{O}')$ in this incomplete representation, the relation $H_K^{\mathcal{O}'}(p) \approx H_K^{\mathcal{O}}(p)$ will hold for modest degrees of contour deletion because the ratio of the number of contour points in each bin to the total number of points will be approximately the same for the complete and the incomplete contour representations. Hence, $d^{SC}(S_{\mathcal{O}}^{SC}, S_{\mathcal{O}'}^{SC}) \approx 0$ for the normalized shape contexts.

We performed experiments with and without the above mentioned normalization of the shape context, Fig. 9. There is a significant performance improvement in the segment-wise deletion and the depletion tests due to the normalization procedure. This justifies the use of this procedure in the experiments whose results are shown in Fig. 6.

4 Further Aspects

4.1 Choice of Data Set

The experiments presented in Section 3 were carried out in the well-known MPEG-7 data set that is a de facto standard for comparison of shape recognition algorithms which use complete representations. As the scope of this paper is to introduce a new test, it is important to check if the conclusions drawn from the ICR test are consistent across data sets. Furthermore, the MPEG-7 data set has a certain restriction:

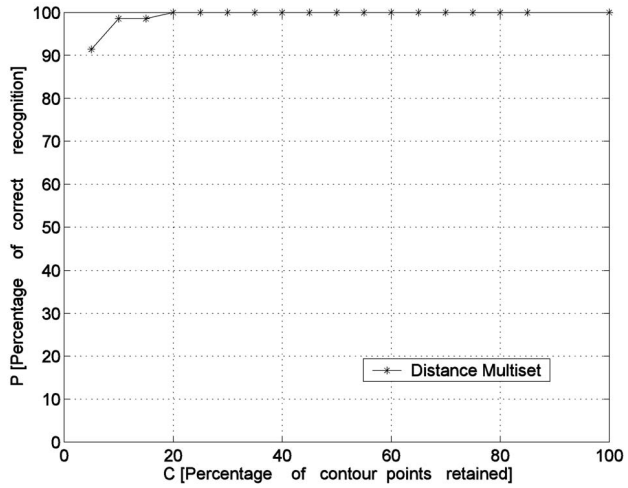


Fig. 8. Performance of the distance multiset algorithm in the segment-wise deletion test with incomplete representations obtained from affine transformed versions of the reference objects. Affine transformations have no substantial effect on the performance of the distance multiset method (compare with Fig. 6a).

The contours that define the objects are only outer contours. In contrast, the contour images used in the original psychophysical test of Gollin (Fig. 2) as well as the contour images provided by computer algorithms (Fig. 3) include inner edges next to the occluding boundaries of the objects.

For these reasons, we carried out experiments on a second data set, the Columbia University Image Library (COIL-20) data set. It contains 1,440 different images divided into 20 classes, each of 72 similar objects. To prepare data for the ICR test, we follow the same procedure as in the case of the MPEG-7 data set. We choose one object from each class (Fig. 10, Row 1) and extract the contours. The resulting 20 contour images are rescaled to a diameter of 76 pixel units, cf., Row 2 of Fig. 10. These 20 rescaled contour images are considered as the complete representations and are used as reference images in our experiments.

We performed the ICR test with these images and the results are shown in Fig. 11. Comparing these results with the results in Fig. 6, we see that there is no qualitative difference in the performances of the algorithms across the data sets.

4.2 Combination of Variability Types

Object recognition methods can be evaluated for their robustness in various respects, viz. affine transformation, variation of shape, presence of noise, etc. In this paper, we put forward a new attribute, namely, robustness to incomplete contour representations. Since this is the focus of this paper, the data set used in the experiments includes only variability regarding contour incompleteness. This type of variability is not combined with other types, e.g., variation of shape or size. The rationale behind this decision is that, first, the algorithms should be characterized by their robustness to one type of variability at a time, e.g., shape changes, incompleteness of contours, rotation, size, etc. Only after such characterization, it makes sense to combine different types of variation, e.g., contour incompleteness with variation in shape or incompleteness with variation in size or any combination of these or other types of variation.

In this context, we note that good performance in the original ICR test does not guarantee good performance in other respects, e.g., robustness to shape or size variation.

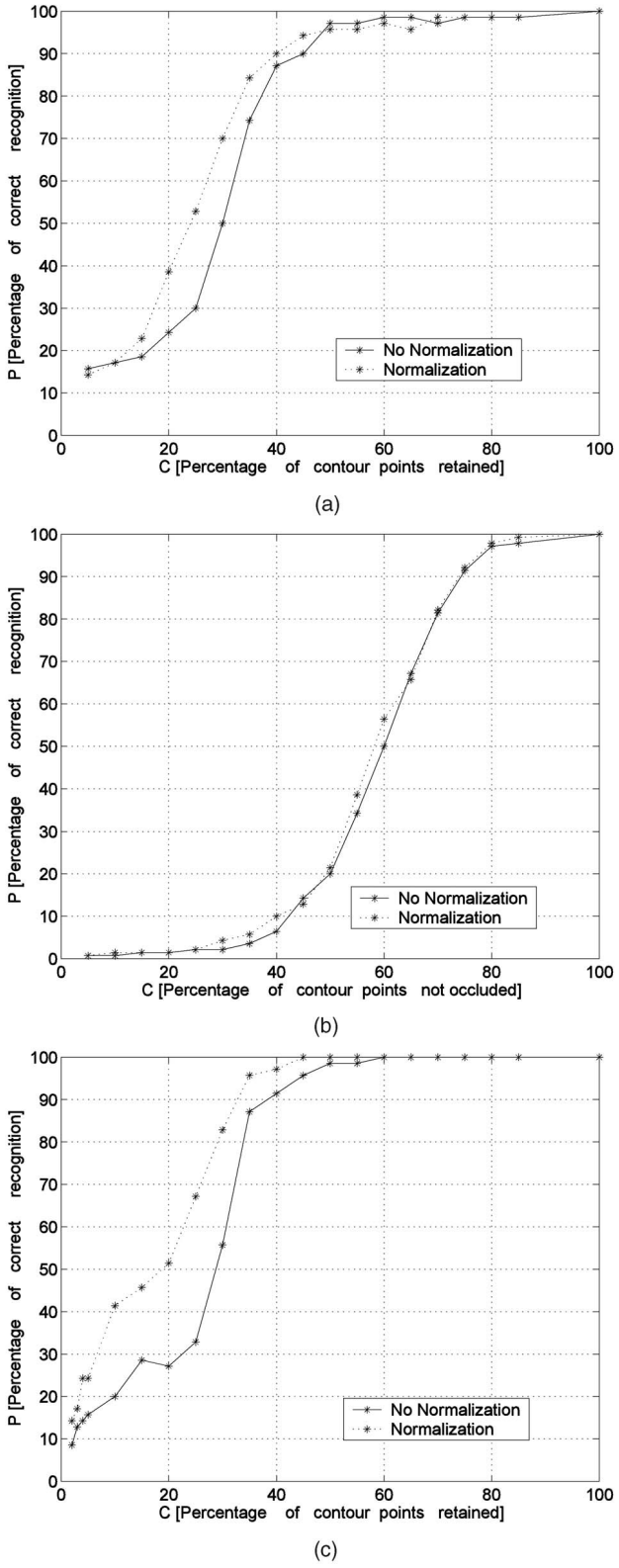


Fig. 9. The performance improvement of the shape context method due to normalization is (a) significant in the segment-wise deletion ICR test, (b) slight in the occlusion ICR test, and (c) large in the depletion ICR test.

Hence, a good performance in the ICR test should be considered as a *necessary condition* for object recognition methods to perform well in a real-world scenario but not as a sufficient one. We are not aware of any evaluation procedure

Fig. 10. Row 1: Sample of the COIL-20 database images used. Row 2: Rescaled contour images that are considered as complete representations.

for shape recognition methods which is sufficient in such respect. Once an algorithm is tested for its robustness to shape variation (e.g., by the MPEG-7 bull's eye test [22], [13]), incompleteness (e.g., by the proposed ICR test), and other types of simple variation, more elaborate tests that combine different types of variation can be applied. Below, we give an example.

Shape variability and incompleteness: To assess robustness to incomplete representations along with robustness to variation in shape, we propose a *bull's eye* ICR test using the MPEG-7 data set. Unlike the basic ICR test proposed above, where we choose one complete contour image from each class of objects, in the bull's eye ICR test we choose more than one complete contour images, say, n (out of 20 available in the MPEG-7 data set), to represent a class. For a given percentage of contour retention c, each of the $70 \times n$ incomplete contour images constructed by the methods described in Section 3 is compared with all $70 \times n$ reference (complete contour) objects and the corresponding n nearest neighbors are found.[4] Let ϵ_i ($\epsilon_i \leq n$) be the number of nearest neighbors that belong to the class of object i. The percentage of correct recognition $P(c)$ is calculated in the following way:

$$P(c) = \frac{\sum_{i=1}^{70\times n} \epsilon_i}{70 \times n^2} \times 100. \quad (14)$$

In Fig. 12, we present the results of a bull's eye depletion test with $n = 2$. Increasing the value of n, we can introduce more shape variation in this test procedure. An extreme case would be to consider the full MPEG-7 data set ($n = 20$).

4.3 Object Size

The object size can have effect on the results of an ICR test through 1) the resolution of the reference objects and 2) a possible mismatch between the size of a reference object and a test object.

Regarding the resolution of the reference objects in our experiments, we found that, for a given percentage of contour degradation (by any method), the performance of the algorithms grows with the diameter of the reference objects. This is in agreement with the results of psychophysical studies on humans where performance increases with the visual angle at which the objects are presented [23]. In Fig. 13, we illustrate the performance of the shape context method in the depletion test for two different sizes of the reference

4. In the original bull's eye test [22], a fixed number of 40 nearest neighbors is used. Here, we use a modification of the test in which the number of nearest neighbors is taken to be equal to the number n of objects in one class.

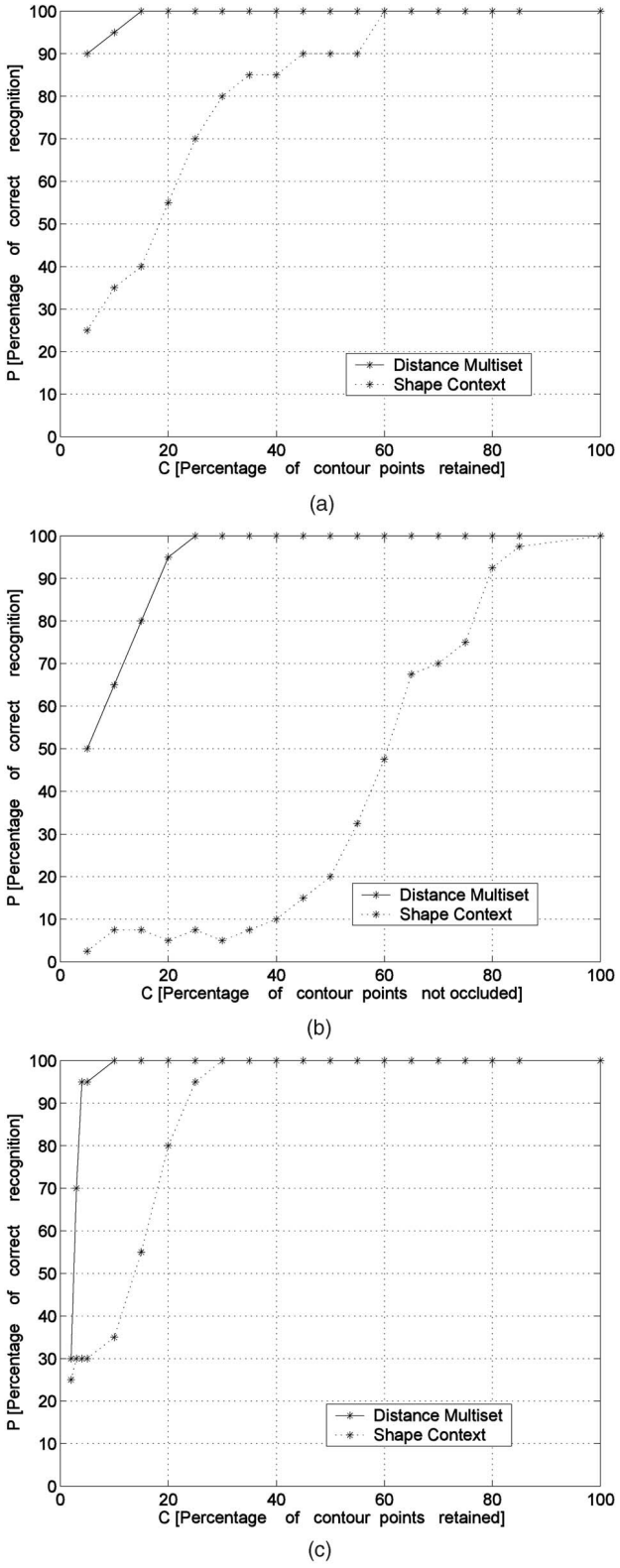


Fig. 11. Results of the ICR tests with a subset of the COIL-20 data set. There is no qualitative difference in performance of the algorithms in the COIL-20 data set compared to the MPEG-7 data set (Fig. 6). (a) Segment-wise deletion ICR test (COIL-20). (b) Occlusion ICR test (COIL-20). (c) Depletion ICR test (COIL-20).

objects. From this figure, we see that, for a low percentage (less than 45 percent) of retained contour the performance of the method is appreciably better in the case when the

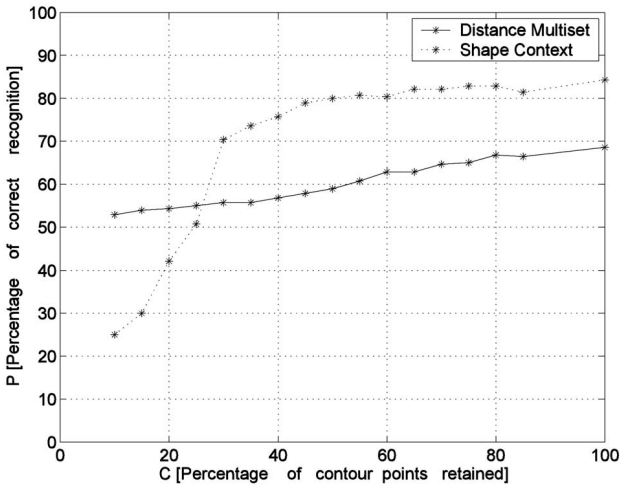


Fig. 12. Results of a bull's eye depletion test, using a subset of 70×2 images of the MPEG-7 data set. For the high percentage of pixels retained, the shape context method performs better than the distance multiset method.

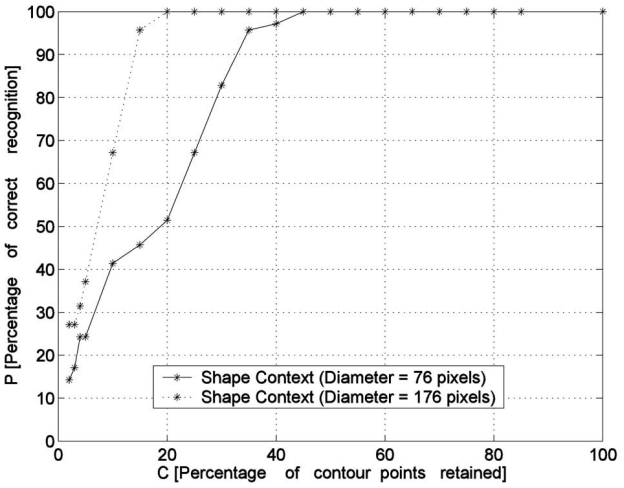


Fig. 13. The performance of the shape context method in the depletion test with two different sizes of the reference objects—the larger the diameter, the better the performance is.

reference object size is bigger. To eliminate this effect and to standardize the test procedure, we rescaled the reference contour images to a fixed diameter (76 pixel units).

The problem of a possible mismatch between the sizes of reference and test objects is not specific for the type of test objects (incomplete contour representations) that we use in this study. The problem is rather related to the way in which shape recognition algorithms deal or do not deal with size variation. Both algorithms (shape context and distance multiset) used here to illustrate the ICR test are not intrinsically scale invariant. In [20], the authors of the shape context method suggest normalization of all radial distances by the mean distance between all point pairs in a contour in order to make the shape context descriptor scale invariant. To achieve the same goal, the authors of the distance multisets method [13] prescribe dividing all distances by the diameter of the object under consideration.

In the ICR test, the reference contour images are rescaled to have the same diameter (76 pixel units) and the incomplete

representations are constructed from these rescaled versions of the reference contour images by removing contour points. Hence, the actual distances between retained points do not change and the algorithms are provided with test images having the same point-to-point distance as the reference images. The distance multiset algorithm benefits slightly more from this aspect of the construction of incomplete representations because the distance multiset of a point from an incomplete contour is a subset of the distance multiset of the same point in the corresponding complete contour. As pointed out above, the dissimilarity computed for pairs of such points will be zero for the distance multiset algorithm. Zero dissimilarity cannot be guaranteed for the shape context method.

The above choice of a procedure for the construction of incomplete representations was made deliberately: We want to quantify the success or failure of an algorithm in coping with incompleteness and, for this purpose, we want to minimize the effects of other aspects, e.g., size variation. In a real-world situation such as the one illustrated by Fig. 3, however, there is no guarantee that the size of a test object with an incomplete contour will match exactly the size of the corresponding reference object. Under such circumstances, it is important that a method can automatically determine the appropriate size. Our experiments can easily be modified for this purpose by rescaling all test images (incomplete contour representations) to the same object diameter (of 76 pixel units) as the reference objects. Fig. 14 shows the results of ICR tests with incomplete representations that have been obtained in this way. A comparison with Fig. 6 shows that performance degrades due to the fact that the sizes of incomplete representations do not exactly match the sizes of the corresponding complete representations. When an incomplete representation is constructed from a complete representation, one of the points for which the diameter of the object is measured can be removed. The consequence is that the diameter of the incomplete representation will decrease. After rescaling this diameter to the standard size (of 76 pixel units), all pairwise distances in an incomplete representation will increase and become larger than their counterpart distances in the corresponding complete representation. The smaller the percentage of contour retention, the larger this effect and the larger the performance degradation.

This method of rescaling has a devastating effect on the performance of the algorithms with incomplete representations obtained through occlusion (Fig. 14b). For this case and, also, for the general case of objects that are not segmented from their background, one should adopt a different multiscale approach. In real-world situations, such as the one illustrated by Fig. 3, an object is not segmented from its background. In contrast, the very purpose of using a shape descriptor in such a situation is to test whether a given object is present in a complex scene and to separate it from the background. Under such circumstances and without any prior knowledge about the appropriate scale to be used, one can take a multiscale approach: Shape descriptors are computed independently at multiple resolutions and the descriptors computed at each scale are compared with the reference descriptors. The multiscale approach has been advocated for both biological [26] and computer [27] vision.

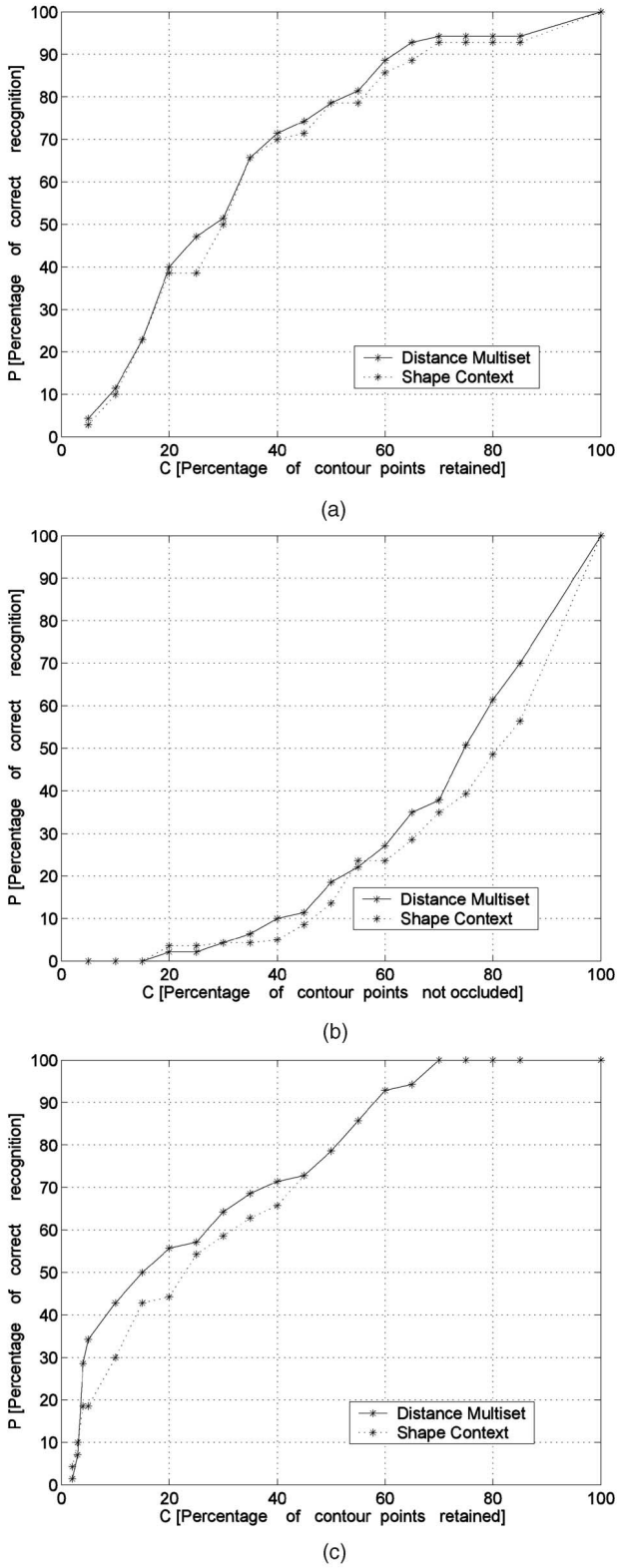


Fig. 14. Results of the ICR tests with incomplete contour representations that are rescaled to a constant diameter of 76 pixel units. The performances of the algorithms are worse than in the case when the distances in the complete and incomplete representations are equal (compare with Fig. 6). The effect is particularly strong for the distance multiset algorithm. (a) Segment-wise deletion test. (b) Occlusion test. (c) Depletion test.

4.4 Criterion for Acceptable Performance

The performance curves obtained in ICR tests can be used to compare algorithms, as illustrated in Fig. 6. It would be interesting to define a criterion for acceptable performance of an algorithm without having to compare it with another algorithm. One possible way of achieving this is to use as a reference the performance of humans in a similar experimental setup. As a matter of fact, similar studies exist in psychophysics [23], [24], [25]. For instance, in [23], the performance of humans in a test that is similar to the segment-wise deletion ICR test is studied whereby identical gaps between fragments of equal length are used. Similar to computer algorithms, the performance of humans depends on the size of the objects. To make a comparison possible, objects need, therefore, to be presented at a certain standard size (visual angle) that is related to the standard object size (in pixel units) deployed in computer algorithms. To establish such a relation, one should use visual acuity data (minimum visual angle between two distinguishable points).

5 Summary and Conclusion

Object recognition methods that employ shape descriptors have been evaluated and compared using various characteristics like invariance, uniqueness, and stability [12]. Marr and Nishihara [28] proposed three criteria for judging the effectiveness of a shape descriptor, viz. accessibility, scope and uniqueness, stability, and sensitivity. Brady [29] put forward a set of criteria for representation of shape, viz. rich local support, smooth extension, and propagation. A detailed survey and comparison of shape analysis techniques on the basis of some of the above mentioned criteria can be found in [14]. In the current work, motivated by characteristics of the human visual system [1], we propose an additional new criterion, viz. robustness to contour incompleteness to compare and characterize contour-based shape recognition algorithms using their performance in recognizing objects with incomplete contours. We are not aware of any such comparison and characterization in the present literature.

We put forward the following procedure, which we call the ICR test:

1. Take a set of images of objects and extract contours. Rescale all contour images to the same object diameter.
2. Train the recognition system with these complete contour representations.
3. Construct different sets of incomplete representations from the complete contour representations, quantifying the level of incompleteness using the percentage of contour pixels retained.
4. Using the incomplete representations as a test set, evaluate the recognition rate as a function of the percentage of contour pixels retained.

We distinguish three different types of incomplete contour representations according to the method used to remove parts of the contour: segment-wise deletion, occlusion, and random pixel depletion. We created test data sets of such incomplete contour representations derived from images from the MPEG-7 and COIL-20 data sets and made them publicly available at www.cs.rug.nl/~petkov/.

We illustrated the test framework with two shape recognition methods based on the shape context and the distance multiset. We should note that other shape recognition methods, such as those based on the Hausdorff measure [30], [31], wavelet descriptors [32], dynamic programming [33], graph matching [34], curve alignment [35], Fourier descriptors [36], can also be studied in this framework. As the main objective of the research presented in this paper is to introduce a new test framework, an exhaustive comparison of different methods under this framework is beyond the scope of this paper. The two methods tested were chosen merely for illustrative purposes and we did not aim to prove the superiority of any method. A complete comparative study of the two methods is out of the scope of this work. In our illustrative experiments, we found that: 1) The distance multiset shape recognition method outperforms the shape context method regarding robustness to contour incompleteness, especially for high levels of incompleteness. 2) Both methods perform similarly to the human visual system in the sense that their performances are increasing functions of the degree of contour completeness and are best in the case of the depletion test and worst in the case of the occlusion test.

Our main conclusions are as follows: The robustness of contour-based shape recognition methods to incompleteness of contour representations is an important aspect of any contour-based objects recognition system. The ICR test as defined and proposed in this paper is an adequate framework for assessing the above mentioned performance and can be used as a standard test procedure for any contour-based object recognition system/algorithm.

Appendix A
Cost of Matching Two Distance Multisets

Consider the multisets

$$X = (x_1, x_2, \ldots x_M), \tag{15}$$

$$Y = (y_1, y_2, \ldots, y_N), \tag{16}$$

where $M \leq N$. Let π be a one-to-one mapping from the set $\{1, \ldots, M\}$ to the set $\{1, \ldots, N\}$ and let Π be the set of all such mappings. The mapping π defines an assignment of a unique element $y_{\pi(i)} \in Y$ to each element $x_i \in X$. The cost $c_\pi(X, Y)$ of a mapping/assignment $\pi \in \Pi$ is defined as follows:

$$c_\pi(X, Y) = \sum_{i=1}^{M} |x_i - y_{\pi(i)}|. \tag{17}$$

Let c be the minimum of the costs of all such possible mappings:

$$c(X, Y) = min\{c_\pi(X, Y) | \pi \in \Pi\}. \tag{18}$$

Note that X and Y are sorted in ascending order by the definition of a distance multiset. To compute $c(X, Y)$

efficiently, we use the algorithm described in [37], which has complexity $O(M(N-M))$.

APPENDIX B
PROOF OF THE LEMMA

Proof. *Claim 1*: The distance multisets of A and B are identical.

By definition, f is an isometry, which implies that distance multisets of A and B are identical, that is, for every $p \in A \; \exists$ a $q \in B$, $q = f(p)$ such that $D_N^A(p) = D_N^B(q)$, assuming that $\mathrm{card}(A) = \mathrm{card}(B) = N$.

Claim 2: $d^{DM}(S_C^{DM}, S_B^{DM}) = 0$.

The definition of the distance multiset along with (18) implies that, for every $p_i \in C$, $\exists$ a $q_j \in B$ such that $c_{i,j}^{DM} = 0$. Hence, the minimum of every row of the cost-matrix of point-wise dissimilarities is 0, which implies, by (9), $d^{DM}(S_C^{DM}, S_B^{DM}) = 0$.

Claim 2 and the invariance of distance multisets in Claim 1 imply that

$$d^{DM}\left(S_C^{DM}, S_A^{DM}\right) = 0.$$

□

REFERENCES

[1] E.S. Gollin, "Developmental Studies of Visual Recognition of Incomplete Objects," *Perceptual and Motor Skills,* vol. 11, pp. 289-298, 1960.

[2] C. Grigorescu, N. Petkov, and M.A. Westenberg, "Contour and Boundary Detection Improved by Surround Suppression of Texture Edges," *Image Vision and Computing,* vol. 22, pp. 609-622, 2004.

[3] C. Grigorescu, N. Petkov, and M. Westenberg, "Contour Detection Based on Nonclassical Receptive Field Inhibition," *IEEE Trans. Image Processing,* vol. 12, pp. 729-739, July 2003.

[4] D.M. Gavrila, "Multi-Feature Hierarchical Template Matching Using Distance Transforms," *Proc. Int'l Conf. Pattern Recognition,* pp. 439-444, 1998.

[5] A. Goshtaby, "Description and Discrimination of Planar Shapes Using Shape Matrix," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 7, pp. 738-743, Nov. 1985.

[6] L. Davis, "Two-Dimensional Shape Representation," *Handbook of Pattern Recognition and Image Processing,* I. Young and K.S. Fu, eds. pp. 233-245, Academic Press, 1986.

[7] H. Blum, "A Transformation for Extracting New Descriptors of Shape," *Models for the Perception of Speech and Visual Forms,* Whaten-Dunn, ed. pp. 362-380, MIT Press, 1967.

[8] S. Peleg and A. Rosenfeld, "A Min-Max Medial Axis Transformation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 3, pp. 208-210, 1981.

[9] S.O. Belkasim, M. Shridhar, and M Ahmadi, "Pattern Recognition with Moment Invariants: A Comparative Study and New Results," *Pattern Recognition,* vol. 24, pp. 1117-1138, 1991.

[10] R.J. Prokop and A.P. Reeves, "A Survey of Moment-Based Techniques for Unoccluded Object Representation and Recognition," *CVGIP: Graphical Models and Image Processing,* vol. 54, pp. 438-460, 1992.

[11] L.J. Latecki and R. Lakämper, "Shape Similarity Measure Based on Correspondence of Visual Parts," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, pp. 1185-1190, 2000.

[12] F. Mokhtarian and A.K. Mackworth, "A Theory of Multiscale, Curvature-Based Shape Representation for Planar Curves," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 8, pp. 789-805, Aug. 1992.

[13] C. Grigorescu and N. Petkov, "Distance Sets for Shape Filters and Shape Recognition," *IEEE Trans. Image Processing,* vol. 12, no. 10, pp. 1274-1286, 2003.

[14] S. Loncaric, "A Survey of Shape Analysis Techniques," *Pattern Recognition,* vol. 31, no. 8, pp. 983-1001, 1998.

[15] R.C. Veltkamp and M. Hagedoorn, "State of the Art in Shape Matching," Technical Report UU-CS-1999-27, Utrecht Univ., 1999.

[16] G. Nagy, "Twenty Years of Document Image Analysis in PAMI," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 1, pp. 38-62, Jan. 2000.

[17] T. Pavlidis, "Algorithms for Shape Analysis of Contours and Waveforms," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 2, pp. 301-312, 1980.

[18] Y. Aloimonos, "Visual Shape Computing," *Proc. IEEE,* vol. 76, pp. 899-916, 1988.

[19] R. Basri, L. Costa, D. Geiger, and D. Jacobs, "Determining the Similarity of Deformable Shapes," *Vision Research,* vol. 38, pp. 2365-2385, 1998.

[20] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 4, pp. 509-522, Apr. 2002.

[21] C. Papadimitroiou and K. Stieglitz, *Combinatorial Optimization.* Englewood Cliffs, N.J.: Prentice-Hall, 1982.

[22] L.J. Latecki, R. Lakämper, and U. Eckhardt, "Shape Descriptors for Non-Rigid Shapes with Single Closed Contour," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 424-429, 1998.

[23] V. Chihman, V. Bondarko, Y. Shelpin, and M. Danilova, "Fragmental Figure Perception," *Perception,* vol. 33, supplement, p. 76a, 2004.

[24] Y. Shelepin, O. Vahromeeva, A. Harauzov, S. Pronin, N. Foreman, and V. Chihman, "Recognition of Incomplete Contour and Half-Tone Figures," *Perception,* vol. 33, supplement, p. 85c, 2004.

[25] N.P. Foreman and R. Hemmings, "The Gollin Incomplete Figures Test: A Flexible, Computerized Version," *Perception,* vol. 16, pp. 543-548, 1987.

[26] J.J Koenderink and A.J. van Doorn, "Visual Detection of Spatial Contrast; Influence of Location in the Visual Field, Target Extent and Illuminance Level," *Biological Cybernetics,* pp. 157-167, 1978.

[27] P.J. Burt, "Smart Sensing with a Pyramid Vision Machine," *Proc. IEEE,* vol. 76, no. 8, pp. 1006-1015, 1988.

[28] D. Marr and H.K. Nishihara, "Representation and Recognition of the Spatial Organization of Three Dimensional Shapes," *Proc. Royal Soc. London, B,* vol. 200, pp. 269-294, 1978.

[29] M. Brady, "Criteria for Representations of Shape," *Human and Machine Vision,* J. Beck, B. Hope, and A. Rosenfeld, eds., pp. 39-84, Academic Press, 1983.

[30] D. Huttenlocher, G. Klanderman, and W. Rucklidge, "Comparing Images Using Hausdorff Distance," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 9, pp. 850-863, Sept. 1993.

[31] D. Huttenlocher, H. Lilian, and C. Olson, "View-Based Recognition Using an Eigenspace Approximation of Hausdorff Measure," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 9, pp. 951-955, Sept. 1999.

[32] G.C.-H. Chuang and J. Kuo, "Wavelet Descriptor of Planar Curves: Theory and Application," *IEEE Trans. Image Processing,* vol. 5, pp. 56-70, Jan. 1996.

[33] E.G.M. Petrakis, A Diplaros, and E. Milios, "Matching and Retrieval of Distorted and Occluded Shapes Using Dynamic Programming," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 11, pp. 1501-1516, Nov. 2002.

[34] A.D.J. Cross and E.R. Hancock, "Graph Matching with a Dual-Step EM Algorithm," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 11, pp. 1236-1252, Nov. 1998.

[35] T.B. Sebastian, P.N. Klein, and B.B. Kimia, "On Aligning Curves," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 1, pp. 116-125, Jan. 2003.

[36] L. Bartolini, P. Ciaccia, and M. Patella, "WARP: Accurate Retrieval of Shapes Using Phase of Fourier Descriptors and Time Warping Distance," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, no. 1, pp. 142-147, Jan. 2005.

[37] N. Petkov, "Algorithm for the Cost of an Optimal Assignment of Two Sets of Real Numbers," Technical Report, 2003-9-07, Inst. of Math. and Computing Science, Univ. of Groningen, 2003.

Anarta Ghosh received the BSc degree (1995) with honors in physics (first class) from Calcutta University, India, ranking first in his college. He received the ME degree (2000) (first class) in electrical engineering from the Indian Institute of Science, Bangalore, ranking second in his class. After working as a digital signal processing engineer for Lucent Technologies (Agere systems) for 10 months and as a visiting scientist at the Fraunhofer Institute for Industrial Mathematics (Kaiserslautern, Germany) for six months, he has been working as a PhD student at the Institute of Mathematics and Computing Science, University of Groningen, The Netherlands, since August 2002. His research interests are in the area of cognitive computer vision, pattern recognition, machine learning, theoretical neuroscience, signal processing, and multimedia systems.

Nicolai Petkov is scientific director (head) of the Research Institute of Mathematics and Computing Science, University of Groningen, The Netherlands, where he also holds the Chair of Intelligent Systems. He is the author of two books and 90 scientific publications. His current research interests are in the area of computer simulations of the visual system, making links between computer vision, neurophysiology, psychophysics, and arts.